UNIVERSITY of York

Discussion Papers in Economics

No. 21/04

Reduced-Form Allocations for Multiple Indivisible Objects under Constraints

Xu Lang, School of Economics, Southwestern University of Finance and Economics,Chengdu, China. Email: langxu@swufe.edu.cn
and
Zaifu Yang, Department of Economics and Related Studies, University of York, York, UK. Email: zaifu.yang@york.ac.uk

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Reduced-Form Allocations for Multiple Indivisible Objects under Constraints*

Xu Lang[†] Zaifu Yang[‡]

July 14, 2021

Abstract

We examine the implementation of reduced-form allocation rules that assign multiple indivisible objects to many agents, with incomplete information and distributional constraints across objects and agents. To obtain implementability results, we adopt a lift-and-project approach, which reduces the problem to a problem of enumerating finite generators of a projection cone. We study geometric and combinatorial properties of the projection cone and provide a total unimodularity condition that leads to several characterization results including those on hierarchies and bihierarchies. Our results have applications in matching markets with constraints where agents may have ordinal or cardinal preferences.

JEL classification codes: D44, C65.

Keywords: Implementation, Reduced-form rules, Indivisible goods, Distributional constraints, Total unimodularity, Incomplete information.

---

*We are very grateful to Jimmy Chan, Eric van Damme, Dolf Talman, and many participants for their helpful feedback on earlier versions of this paper in seminars, workshops and conferences. Any remaining errors are, however, our own responsibility.

[†]School of Economics, Southwestern University of Finance and Economics, Chengdu, China. Email: *langxu@swufe.edu.cn*

[‡]Department of Economics and Related Studies, University of York, York, UK. Email: *zaifu.yang@york.ac.uk*

1 Introduction

This paper aims to study an implementation problem of reduced-form allocation rules for assigning multiple indivisible objects to many agents. A salient feature of the problem is that it can accommodate a variety of distributional constraints across objects and agents, including typical floor and ceiling constraints in environments of incomplete information. The reduced-form approach goes back to Myerson (1981) on auction design. In his paper, the seller's problem is expressed as a revenue maximization over the set of feasible and incentive compatible allocation rules and transfers. A buyer's incentive constraint is then used to express his transfers in terms of interim allocation probabilities, which reduce the problem to an optimization over interim allocation probabilities only, i.e., the *reduced-form*. To apply this approach, one should be able to describe the set of feasible interim allocation rules (or reduced forms). In single-unit auctions, Maskin and Riley (1984) and Matthews (1984) first study the implementability condition, and Border (1991) derives a characterization, nowadays known as Border's theorem.

For a multi-unit auction model with capacity constraints, Che et al. (2013) develop a network flow method to characterize the implementability condition. In their method, the implementation problem is first transformed into a feasible network flow problem and then existence results from graph theory are invoked to obtain their characterization results. Goeree and Kushnir (2016) examine reduced-form implementation for environments with interdependent valuations. Alternative characterizations have also been found. Manelli and Vincent (2010) and Gershkov et al. (2013) observe an important equivalence of Bayesian and dominant strategy implementation. Hart and Reny (2017) obtain a related characterization. Alaei et al. (2019) propose a polymatroidal decomposition method and obtain their characterization results. Meanwhile, Gopalan et al. (2015), Cai et al. (2012), and Alaei et al. (2019) examine the computational complexity of the reduced-form approach. A common feature of these models is that there are no side constraints across different items.

In this paper we investigate the reduced-form allocations of multiple indivisible items to many agents with a variety of distributional constraints, going beyond those traditional ones mentioned above. We briefly discuss several important cases of distributional constraints that our model intend to cover. In many practical situations, markets are regulated by distributional policies or constraints such that allocations across different objects are not independent. A typical example is the "regional cap" in Japanese residency matching program (Kamada and Kojima, 2015) that matches hospitals (agents) with doctors (objects). To regulate the geographical distribution of doctors, the total number of the doctors matched within a region is subject to a "regional cap". Another important example is college admission. Higher education institutions usually set lower quotas for each of their particular areas of study. If the number of assigned students is less than this quota for a particular area, then the project has to be cancelled for that year (Biró et al., 2010, Ehlers et al., 2014, Fragiadakis and Troyan, 2016). Course allocation is yet another example (Budish and Cantillon, 2014, Budish et al., 2013, Sönmez and Ünver, 2010). In a university department, students seek to take multiple courses as part of their programmes. Each student can take at most one seat in each course. For each course, there are ceiling and floor on the number of seats. In addition, an aggregate capacity constraint may restrict the total number of seats for the courses within the department. In China, every region requires a certain minimum number of doctors as a floor constraint in its area hospitals and hospitals also face hierarchical constraints of recruiting doctors (Cheng and Yang, 2017). The presence of these distributional constraints poses a challenge for reduced-form allocation implementation: The problem cannot be treated as separate single-object problems and the reduced form for every agent is multidimensional.

To be precise, the problem under consideration concerns the allocation of a finite set of heterogeneous indivisible objects to a finite set of agents subject to distributional constraints. Objects can be private goods which will be consumed privately and independently by agents. Objects can be public or club goods like courses shared by students. Every agent may demand several objects and is associated with a finite set of types representing her private information about her preferences. There will be distributional constraints across different combinations of agents and objects. Constraints include floor and ceiling constraints across both agents and objects. We use a lift-and-project approach to obtain a characterization of the implementability condition. This approach was first introduced in polyhedral combinatorics (Balas, 2001, Balas and Pulleyblank, 1983) and later used by Vohra (2013) to study linear characterizations for combinatorial objects, including reduced-form auctions. Briefly speaking, in this approach, by lifting, the combinatorial object of interest (i.e., reduced forms) is first formulated by a linear system in some higher-dimensional space. Then by constructing a projection cone and finding its finite generators, it gives arise to the linear system of interest.

To obtain a complete description of the generators, we investigate geometric and combinatorial properties of the projection cone. We provide a general sufficient condition on the projection cone such that an enumeration of the generators is possible. The general condition underlying our characterization is called *total unimodularity*. It captures the essential feature of a class of well-behaved totally unimodular matrices extensively studied in discrete mathematics (Schrijver, 1986). It is also well-known that there exists a polynomial time recognition algorithm for totally unimodular matrices; see Seymour (1980, 1981) and Schrijver (1986). In other words, it is very easy to verify the general condition. We show that if the projection cone preserves total unimodularity, then the generators of the projection cone can be enumerated. We then identify two large classes of constraint structures where the projection cone preserves total unimodularity: *Hierarchies and bihierarchies with canonical constraints*. Hierarchical structures are common in organizations (firm, hospital, or university) and various markets, and have been well-studied in the literature; see e.g., Williamson (1975) and Demange (2004). They are called laminars in mathematics (Fujishige, 2005). Bihierarchies are the union of two disjoint hierarchies as a generalization of hierarchies and are recently investigated by Budish et al. (2013).

We concentrate on the universal implementation where the result does not depend on the specification of quotas. In practice, the designer may have no information on capacity and a universal implementation is therefore very desirable (Budish et al., 2013, Che et al., 2013). Our characterization results are algebraic and very general, covering a variety of distributional constraints such as floor constraints, ceiling constraints, and mixed floor and ceiling constraints. Our first major result (Theorem 1) establishes a general characterization on implementation, showing that *total unimodularity is a sufficient condition*. Our second major result (Theorem 2) proves that *hierarchies and bihierarchies with canonical constraints* suffice to guarantee total unimodularity. Our third and fourth results (Theorems 3 and 4) examine two important special cases: floor constraints and ceiling constraints. We also discuss the relationship between our results and those of Border (1991), Che et al. (2013), Budish et al. (2013), Alaei et al. (2019), and others. Furthermore, we apply our current approach to the compromise model of Börgers and Postl (2009) and the bilateral trade model of Myerson and Satterthwaite (1983).

The rest of this paper is organized as follows. Section 2 introduces a general model on reduced-form implementation. Section 3 introduces our main characterization theorems on implementation. Section 4 discusses the lift-and-project approach and several lemmas which play a key role in proving the characterization theorems. Section 5 studies some special class of problems including a bilateral trade model and a compromise model. Section 7 concludes.

2 The Model

We study a model in which a finite set A of m heterogeneous indivisible objects (e.g. workers, doctors, goods, and courses) is allocated to a finite set N of n different agents (e.g. firms, hospitals, and students) under a variety of constraints. The two numbers n and m may be equal or different. Note that the set A can also accommodate multiple identical objects which will be labelled differently. A **pure outcome** $x = (x(i,j))$ is described as an $n \times m$ matrix indexed by all agents i and objects j, where each entry $x(i,j) \in \{0,1\}$ is the quantity of object j that agent i receives. Note that it is possible for each agent to receive several objects and also for one object like course to be shared by several agents. Generally speaking, in the allocation of private goods, no object will be assigned to more than one agent, while in the case of shared or public goods, every object can be jointly consumed by multiple agents.

A set $G \subseteq N \times A$ of agent-object pairs is called **a constraint set**. Every pair $(i,j) \in N \times A$ is called **a singleton**. For a constraint set G, we define $x(G) = \sum_{(i,j)\in G} x(i,j)$. Each constraint set G is associated with two integer numbers $b(G), c(G) \in \mathbb{Z}_+$ with $b(G) \leq c(G)$ as its floor and ceiling quotas, respectively. A collection of constraint sets $\mathcal{G} \subseteq 2^{N\times A}$ is called **a constraint structure**. Every constraint structure $\mathcal{G}$ is associated with a quotas system (b,c), where we have $(b(G), c(G))$ for every $G \in \mathcal{G}$.

Given a constraint structure $\mathcal{G}$ and its quotas system (b,c), we say a pure outcome

x is **feasible** if

$$b(G) \leq x(G) \leq c(G) \text{ for all } G \in \mathcal{G}. \tag{1}$$

Clearly, for every feasible pure outcome x, we have $0 \leq x(i,j) \leq 1$ for all $(i,j) \in N \times A$ and so $\mathcal{G}$ contains all singletons.

The constraint structure $\mathcal{G}$ with its quotas system (b,c) covers a variety of allocation problems. We briefly discuss two major allocation problems and their implications on their quotas system. For any $j \in A$ and any $G \in \mathcal{G}$, let $G(j) = \{(i,j) \mid (i,j) \in G\}$. In any allocation problem of private goods, usually no item $j \in A$ will be assigned to more than one agent. This means that if x is a feasible pure outcome, for every $G \in \mathcal{G}$ and every $G(j) \in \mathcal{G}$, we must have $0 \leq b(G(j)) \leq x(G(j)) \leq c(G(j)) = 1$. However, for any allocation problem of shared goods like courses or public goods, every object $j \in A$ is typically shared by multiple agents. In this case, for every $G \in \mathcal{G}$ and every $G(j) \in \mathcal{G}$, the quota $c(G(j))$ can be greater than 1.

Two-sided assignment problems often have quotas for every object and quotas for every agent. This corresponds to the classical model of assignment markets; see Crawford and Knoer (1981), Demange and Sotomayor (1986), Koopmans and Beckmann (1957), Shapley and Shubik (1971). Following Budish et al. (2013), we call this kind of constraint structure $\mathcal{G}$ a **canonical two-sided constraint structure**, if $\mathcal{G}$ contains all sets $\{i\} \times A$ for each $i \in N$ (i.e., all rows) and all sets $N \times \{j\}$ for every $j \in A$ (i.e., all columns). This is a special and important class of constraint structures obviously covered by the general framework of constraint structures described above.

Let X denote the set of feasible pure outcomes. A **random outcome** is a matrix $x = (x(i,j))$ indexed by agents and objects where $x(i,j) \in [0,1]$ is a fractional allocation of object $j \in A$ assigned to agent $i \in N$. A random outcome x is **feasible** if it can be described as a lottery over the set of feasible pure outcomes, that is, if there exist nonnegative numbers λ_k summing up to one and feasible pure outcomes $x^k \in X$ such that

$$x = \sum_{x^k \in X} \lambda_k x^k. \tag{2}$$

Let $\Delta(X)$ denote the set of all feasible random outcomes.

Every agent $i \in N$ is associated with a finite set T_i of possible types, which represents agent i's private information about her preference. A type $t_i \in T_i$ may represent agent i's preference ordering $\succsim_i$ over the set A of objects, or it may determine her cardinal utility function, i.e., a payoff vector $v_i \in \mathbb{R}^A$ that assigns a valuation $v_i(j)$ for object j. Hence our model allows for domains that cover ordinal and cardinal preferences. Let $T = \times_{i \in N} T_i$ denote the entire type set, i.e., the product of the type set T_i over all agents $i \in N$, and $T_{-i} = \times_{j \neq i} T_j$. For every $i \in N$, let $\lambda^i : T_i \to \Delta(T_{-i})$ be a belief function, i.e., $\lambda^i(t_{-i}|t_i)$ is the probability that agent i assigns to other agents' type $t_{-i} \in T_{-i}$ when i's type is t_i. We assume that there exists a common prior probability $\lambda \in \Delta(T)$ such that the beliefs of the agents are the posteriors, and $\lambda(t) > 0$ for all $t \in T$. Let λ_i denote i's marginal probability of λ.

We now introduce two large classes of constraint structures. The first one is the class of hierarchies, which has been used by Williamson (1975) and Demange (2004) in different contexts, also called laminars in mathematics (Fujishige, 2005).

Definition 1. *A constraint structure $\mathcal{G}$ is a* ***hierarchy*** *(or a laminar) if for all $G, G' \in \mathcal{G}$,*

$$G \subset G', \ \text{ or } G' \subset G, \ \text{ or } G' \cap G = \emptyset. \tag{3}$$

Our second one concerns a richer and more general class of constraint structures due to Budish et al. (2013), called bihierarchies.

Definition 2. *A constraint structure $\mathcal{G}$ is a* ***bihierarchy*** *if it is the union of two disjoint hierarchies $\mathcal{G}_1$ and $\mathcal{G}_2$, i.e., $\mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2$ and $\mathcal{G}_1 \cap \mathcal{G}_2 = \emptyset$.*

For bihierarchical constraint structures, Budish et al. (2013) obtain a generalization of Birkhoff-von Neumann theorem which is used to characterize their feasible random outcomes; see their Theorem 1.[1] For our current model, we will use bihierarchies as an example of our general conditions to obtain implementation results.

Below we give four examples of different constraint structures. The first two concern bihierarchies and the other two are hierarchies.

[1] Note that the model and objective of Budish et al. (2013) are considerably different from ours. They develop ex ante efficient and fair random allocation mechanisms for environments with various constraints. We deal with a general model with incomplete information and various constraints and investigate interim incentive-compatible allocation mechanisms. In our model every agent $i \in N$ has a finite set T_i of types, while in Budish et al. (2013) incomplete information is not discussed and in their model the set T_i of every agent i could be understood to contain only one element.

Example 1. Suppose that there are three hospitals (agents) $N = \{i_1, i_2, i_3\}$, and three doctors (objects) $A = \{j_1, j_2, j_3\}$. Each doctor can enroll in at most one hospital while each hospital may hire several doctors. The regional cap may require that at most two doctors should be matched by the first two hospitals i_1 and i_2 in the same region. The family of constraint sets is given by (i, j) for every $i \in N$ and every $j \in A$, $N \times \{j\}$ for every $j \in A$, and $\{i_1, i_2\} \times A$, which is a bihierarchy. Each hospital has a set of types that represent its preference over doctors. For instance, let $T_i = \{\succsim_i, \succsim_i'\}$ for some agent i, where

$$\succsim_i: j_1 \succsim_i j_2 \succsim_i j_3, \ \textit{and} \ \succsim_i': j_2 \succsim_i' j_3 \succsim_i' j_1.$$

Let x and x' be two random outcomes with $x(i) = (\frac{1}{2}, \frac{1}{3}, \frac{1}{4})$ and $x'(i) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{3})$. Then $x(i)$ (first order) stochastically dominates $x'(i)$ at $\succsim_i$ in the sense of Bogomolnaia and Moulin (2001). The following linear system describes all possible outcomes:

$$\begin{aligned} 0 \le \textstyle\sum_{i\in N} x(i,j) \le 1 & \quad \text{for every } \ j \in A \\ 0 \le \textstyle\sum_{(i,j)\in\{i_1,i_2\}\times A} x(i,j) \le 2 & \\ 0 \le x(i,j) \le 1 & \quad \text{for every } \ (i,j) \in N \times A \end{aligned}$$

Example 2. Suppose that there are three students $N = \{i_1, i_2, i_3\}$, one compulsory courses c_1, and three optional courses $O = \{o_1, o_2, o_3\}$. Every optional course $o \in O$ faces a floor constraint which requires the course being selected by at least one student for it to open; and every student i is required to take at least one of the optional courses. The family of constraint sets is given by (i, o) for every $i \in N$ and $o \in O \cup \{c_1\}$, $\{i\} \times O$ for every $i \in N$, and $N \times \{o\}$ for every $o \in O \cup \{c_1\}$, which is a bihierarchy, having a canonical two-sided constraint structure. We may assume each student i has a set T_i of types that represent her payoff vectors over optional courses, i.e., $t_i = (1, 4, 2)$. Then with additive valuations, agent i prefers random outcome x to x' at t_i if and only if $x(i) \cdot t_i \ge x'(i) \cdot t_i$. The following system of linear inequalities describes all possible outcomes:

$$\begin{aligned} 1 \le \textstyle\sum_{i\in N} x(i,j) \le \infty & \quad \text{for every } \ j \in O \\ 1 \le \textstyle\sum_{j\in O} x(i,j) \le \infty & \quad \text{for every } \ i \in N \\ 3 \le \textstyle\sum_{i\in N} x(i,c_1) \le 3 & \\ 0 \le x(i,j) \le 1 & \quad \text{for every } \ (i,j) \in N \times (O \cup \{c_1\}) \end{aligned}$$

Example 3. Suppose that there are two students $N = \{i_1, i_2\}$, one compulsory course c_1 and three optional courses $O = \{o_1, o_2, o_3\}$. Every student is required to take c_1, and at least one and at most two of optional courses. Every optional course has no capacity requirement but there is an aggregate capacity $K \in \mathbb{Z}_+$ for all optional courses. The family of constraint sets is given by (i, o) for every $i \in N$ and every $o \in O$, $\{i\} \times O$ for every $i \in N$, and $N \times O$, which is a hierarchy and has a canonical one-sided constraint structure. The following system of linear inequalities describes all possible outcomes:

$$
\begin{aligned}
0 \leq \textstyle\sum_{(i,j)\in N\times O} x(i,j) &\leq K \quad \text{for every } \ j \in O \\
1 \leq \textstyle\sum_{j\in O} x(i,j) &\leq 2 \quad \text{for every } \ i \in N \\
2 \leq \textstyle\sum_{i\in N} x(i,c_1) &\leq 2 \\
0 \leq x(i,j) &\leq 1 \quad \text{for every } \ (i,j) \in N \times (O \cup \{c_1\})
\end{aligned}
$$

Example 4. Suppose that there are three students $N = \{i_1, i_2, i_3\}$ and three fields of studies $A = \{j_1, j_2, j_3\}$ that the students can enroll. Suppose the fields j_1 and j_2 are economics while j_3 is biology. Both i_1 and i_2 have major in economics while i_3 has a different major. Suppose there are no ceiling and floor constraints on the number of seats for each field of studies separately, but there is an aggregate ceiling $L_1(L_2) \in \mathbb{Z}_+$ and floor $B_1(B_2) \in \mathbb{Z}_+$ quotas with $B_1 \leq L_1$ ($B_2 \leq L_2$) on (1) the total number of seats in the first two fields of economics for students major in economics, and (2) the total number of seats in all the fields for all the students. The family of constraint sets is given by (i, j) for every $i \in N$ and every $j \in A$, $\{i_1, i_2\} \times \{j_1, j_2\}$, and $N \times A$, which is a hierarchy but does not have a canonical one-sided constraint structure. The following system of linear inequalities describes all possible outcomes:

$$
\begin{aligned}
B_1 \leq \textstyle\sum_{(i,j)\in\{i_1,i_2\}\times\{j_1,j_2\}} x(i,j) &\leq L_1 \\
B_2 \leq \textstyle\sum_{(i,j)\in N\times A} x(i,j) &\leq L_2 \\
0 \leq x(i,j) &\leq 1 \quad \text{for every } \ (i,j) \in N \times A
\end{aligned}
$$

2.1 Reduced-form implementation

Given any welfare objective at the ex ante or interim stage, we are interested in Bayesian incentive compatible mechanisms that may allow random outcomes. Using the reduced-form approach, we first optimize over the set of incentive compatible interim allocation rules and find an optimal solution. We then ask whether this interim optimal solution can be implemented by an ex post feasible allocation rule or not. The reduced-form implementation problem is to characterize implementable allocation rules based on interim incentive compatibility.

A **feasible ex post allocation rule** $p : T \to [0,1]^{N\times A}$ assigns a feasible random outcome $p(\cdot, t)$ for each type profile $t \in T$, where $p(i, j, t)$ is a fractional quantity of object $j \in A$ assigned to agent $i \in N$. In particular, a mechanism is deterministic if it assigns a pure outcome for each type profile. A feasible ex post allocation rule p induces an interim allocation rule $Q = (Q_i)_{i\in N}$, where $Q_i : T_i \to \mathbb{R}^A$ is agent i's interim expected random allocation. For each $i \in N$, $t_i \in T_i$, and $j \in A$,

$$Q_i(t_i, j) := \sum_{t_{-i}\in T_{-i}} p(i, j, t)\lambda^i(t_{-i}|t_i). \tag{4}$$

Example 5. (Example 1 continued) Suppose for the same agent i, the interim allocation rule is given by $Q_i(\succsim_i) = (\frac{1}{2}, \frac{1}{3}, \frac{1}{3})$ and $Q_i(\succsim_i') = (\frac{1}{3}, \frac{1}{2}, \frac{1}{3})$. Then $Q_i(\succsim_i)$ stochastically dominates $Q_i(\succsim_i')$ at $\succsim_i$, and the dominance is reversed at $\succsim_i'$. Hence Q_i is incentive compatible for agent i.

An interim allocation rule Q is **implementable** if there exists a feasible ex post allocation rule p such that (p, Q) satisfies (4). We then say Q is the **reduced form** of p and p implements Q. Let $\mathcal{Q}$ denote the set of all implementable interim allocation rules. Let $(N, A, T, \lambda, \mathcal{G}, b, c)$ represent **our implementation problem**. We call $(N, A, T, \mathcal{G})$ **the implementation structure.** Define $d := |N||A||T|$ and $l := \sum_{i\in N} |T_i||A|$ and $r := |\mathcal{G}||T|$. Note that the set of feasible ex post allocation rules is defined by a set of linear inequalities and the reduced form operation is a linear map. Hence for the given set $\mathcal{Q}$ the implementable interim rules can be also defined by a set of linear inequalities:

$$\mathcal{Q} = \{Q \in \mathbb{R}^l \mid MQ \leq u\} \tag{5}$$

for some matrix M and vector u. The goal of our implementation problem is to find a linear system of (M, u) that describes the set $\mathcal{Q}$. The system (M, u) is called *a linear characterization* on the set of implementable interim allocation rules.

3 Main Results

In this section, we establish our main characterization theorems for the implementation problem. To obtain these results, we use a lift-and-project approach. This approach starts with a linear system in terms of both ex post and interim allocation rules. By projecting away the variables of ex post allocation rules, we obtain a linear system of interim allocation rules. The procedure then reduces the implementation problem to a problem of enumerating the generators of the projection cone. Since the analysis is involved, we first state our characterization theorems in this section and then discuss the lift-and-project approach in the next section. Most of the proofs will be deferred to the Appendix.

For the convenience of the reader we first review several basic mathematical concepts concerning cones, generators, and extreme rays that will be often used in our analysis. Let $x_1, ..., x_k \in \mathbb{R}^q$ be given vectors. A linear combination $\alpha_1 x_1 + ... + \alpha_k x_k$ is *conic* if $\alpha_1, ..., \alpha_k \geq 0$. The *cone* generated by a finite set $X \subset \mathbb{R}^q$ is the set of all conic combinations of the elements from X, denoted by cone(X), and we call the vectors in X the *generators*. A cone P is polyhedral if $P = \{x \in \mathbb{R}^q \mid Mx \leq 0\}$ for some matrix M. We say P is *pointed* if $Mx = 0$ implies $x = 0$. Otherwise, we say P is *non-pointed*. A nonzero element x of a pointed cone P is called *an extreme ray* if there are $q - 1$ linearly independent constraints binding at x. For detail we refer to Schrijver (1986).

We first consider the implementation problem $(N, A, T, \lambda, \mathcal{G}, b, c)$ with general constraint structures. To proceed, we introduce the following two $\{0, 1\}$ matrices B and C which will be central for our analysis.

- For every $p \in \mathbb{R}^d$, $Q \in \mathbb{R}^l$ and $b, c \in \mathbb{R}^r$, we define the corresponding probability weighted variables x, y, b^λ, and c^λ by multiplying each $p(i, j, t), b(G), c(G)$ by $\lambda(t)$, and each $Q_i(t_i, j)$ by $\lambda_i(t_i)$.

- Matrix C: An incidence matrix where each row is indexed by (G, t), each column is indexed by (i, j, t), and the entry in row (G, t) and column (i, j, t') is 1 if $(i, j) \in G$ and $t = t'$, and 0 otherwise.

- Matrix B: An incidence matrix where each row is indexed by (i, t_i, j), each column is indexed by (i, j, t), and the entry in row (i, t_i, j) and column (i', j', t') is 1 if $i = i'$, $j = j'$, and $t_i = t'_i$, and 0 otherwise.

We use Example 2 to illustrate how the matrices B and C are constructed. For any positive integer k, let I_k denote the identity matrix of order k. For the purpose of illustration, we assume that there are two students, two optional courses, and every student has two types.

Example 6. (Example 2 continued) *The constraint matrices B and C are given by* $B = \begin{bmatrix} B_1 & \\ & B_2 \end{bmatrix}$, $C = \begin{bmatrix} C_{d,1} & C_{d,2} \\ C_{s,1} & \\ & C_{s,2} \\ & I_{16} \end{bmatrix}$, *where* $B_1 = B_2 = [b_{k,l}]_{4\times 8}$ *with* $b_{k,l} = 1$ *if* $k = (i, t_i, j)$ *and* $l = (i, j, t)$, *and* $b_{k,l} = 0$ *otherwise, and* $C_{d,1} = C_{d,2} = I_8, C_{s,1} = C_{s,2} = \begin{bmatrix} I_4 & I_4 \end{bmatrix}$. *Here each B_k denotes the constraint matrix of the reduction form equalities for each course O_k, and C_d and C_s denote the constraint matrices of the canonical row and column constraints, and I_{16} denotes the constraint matrix for singletons.*

The implementation system associated with the matrices B and C arising from the implementation problem described above is given by

$$F = \{(x, y) \in \mathbb{R}^{d+l} \mid b^\lambda \leq Cx \leq c^\lambda, y = Bx\}. \tag{6}$$

We define the **projection cone** by

$$W = \{(f, g, h) \in \mathbb{R}^l \times \mathbb{R}^{2r} \mid -B^\top f + C^\top g - C^\top h = 0, g \geq 0, h \geq 0\}. \tag{7}$$

The constraint matrix of the projection cone W is given by

$$M(W) = \begin{bmatrix} -B^\top & C^\top \\ O & I \end{bmatrix}. \tag{8}$$

Remark 1. The projection cone W depends on the implementation structure $(N, A, T, \mathcal{G})$ but not on the quotas and the beliefs. It implies that the same set of generators would arise if two problems differ only in the quotas and beliefs. Hence the conic approach is, by construction, a universal implementation in the sense of Che et al. (2013) and Budish et al. (2013).

The first main result (i.e., Theorem 1) of the paper will provide a sufficient condition on W such that a complete description of the generators $\hat{W}$ can be found.

The characterization depends on a class of integral matrices called *totally unimodular matrices*. This is a class of well-known and well-studied matrices (Schrijver, 1986). Formally, a matrix M is **totally unimodular** (TUM), if every nonsingular square submatrix has determinant equal to either -1 or $+1$. It is well known that there exists a polynomial time recognition algorithm for totally unimodular matrices; see Seymour (1980, 1981) and Schrijver (1986). In other words, one can easily verify whether a matrix is TUM or not.

Definition 3. *Let $(N, A, T, \mathcal{G})$ be an implementation structure. We say the projection cone W **preserves total unimodularity**, if the constraint matrix $M(W)$ given by (8) is totally unimodular.*

Theorem 1 describes a key characterization of the implementability condition. It shows that if the projection cone preserves total unimodularity, then every generator (up to positive scaling) is a $\{0, \pm 1\}$ vector. That is,

$$(f, g, h) \in \hat{W} \implies (f, g, h) \in \{0, \pm 1\}^l \times \{0, +1\}^{2r}.$$

Hence every f corresponds to a sign function of some sets $S_i^+, S_i^- \subseteq T_i \times A$ for each $i \in N$. In this way we obtain a complete description of the set of implementable reduced forms.

Define an effective bound for any two sets $G, H \subseteq N \times A$ with $G \cap H = \emptyset$ by

$$\beta(G, H) = \max\{x(G) - x(H) \mid x \in \Delta(X)\}. \tag{9}$$

Effective bounds are introduced in Che et al. (2013) and here we extend this idea from their model with multiple units of a single good to the current one with multiple units and multiple goods. A supporting hyperplane argument implies that $P(\beta) = \{x \mid x(G) - x(H) \leq \beta(G, H)$ for all disjoint sets $G, H \subseteq N \times A\}$ completely describes the original feasible set of random outcomes.

Let $\Omega = \cup_{i \in N}(T_i \times A)$. For any $S_i \subseteq T_i \times A$, $i \in N$, let $S = \cup_{i \in N} S_i$. Then $S \subseteq \Omega$. For any $t \in T$ and $S \subseteq \Omega$, we define their intersection by

$$I(t, S) = \{(i, j) \in N \times A \mid (t_i, j) \in S_i\}.$$

We are ready to present the first major characterization result.

Theorem 1. (General Characterization Theorem) *Let* $(N, A, T, \lambda, \mathcal{G}, b, c)$ *be an implementation problem. Suppose the projection cone preserves total unimodularity. Then* $Q \in \mathbb{R}^l$ *is implementable if and only if for all* $S_i^+, S_i^- \subseteq T_i \times A$, $S_i^+ \cap S_i^- = \emptyset$, *for each* $i \in N$,

$$\sum_{i\in N}[\sum_{(t_i,j)\in S_i^+} Q_i(t_i,j)\lambda_i(t_i) - \sum_{(t_i,j)\in S_i^-} Q_i(t_i,j)\lambda_i(t_i)] \leq \sum_{t\in T}\lambda(t)\beta(I(t,S^+), I(t,S^-)). \tag{10}$$

Condition (10) provides a compact description for an implementable reduced form with both the ceiling and floor constraints, as the left hand side may contain both positive and negative entries and the bound β is implicitly determined by the ceiling and floor in condition (9). In general, condition (10) is not separable in S^+ and S^-, since a priori both S^+ and S^- can take non-empty collections of sets. For some cases, Theorem 1 can be further reduced to separate expressions for ceiling constraints and floor constraints by setting $S^- = \emptyset$ or $S^+ = \emptyset$ (Theorems 3 and 4). On the other hand, condition (10) may contain redundant inequalities. We use the following example to illustrate how to obtain condition (10).

Example 7. (Example 3 continued) Suppose for students i_1 and i_2, $T_1 = \{t_1^1, t_1^2\}$ and $T_2 = \{t_2^1, t_2^2\}$. Notice that the compulsory course can be treated separately and we restrict our attention to optional courses. Pick $S_1^+ = \{(t_1^2, o_2)\}$, $S_2^+ = \{(t_2^1, o_1), (t_2^1, o_2), (t_2^1, o_3)\}$, $S_1^- = \{(t_1^1, o_1), (t_1^1, o_2), (t_1^1, o_3)\}$, $S_2^- = \{(t_2^2, o_2)\}$. We calculate β in Table 1. In particular, for (t_1^1, t_2^1) consider the following problem

Table 1: Parameters in Example 7.

t	$I(t,S^+)$	$I(t,S^-)$	β
(t_1^1, t_2^1)	$\{(i_2,o_1),(i_2,o_2),(i_2,o_3)\}$	$\{(i_1,o_1),(i_1,o_2),(i_1,o_3)\}$	$c(\{i_2\}\times O) - b(\{i_1\}\times O)$
(t_1^1, t_2^2)	$\{\emptyset\}$	$\{(i_1,o_1),(i_1,o_2),(i_1,o_3),(i_2,o_2)\}$	$-b(\{i_1\}\times O)$
(t_1^2, t_2^1)	$\{(i_1,o_2),(i_2,o_1),(i_2,o_2),(i_2,o_3)\}$	$\{\emptyset\}$	$c(N\times O)$
(t_1^2, t_2^2)	$\{(i_1,o_2)\}$	$\{(i_2,o_2)\}$	$c(\{i_1\}\times\{o_2\})$

$$\begin{aligned} \max\ \ & x(i_2,o_1) + x(i_2,o_2) + x(i_2,o_3) - x(i_1,o_1) - x(i_1,o_2) - x(i_1,o_3) \\ s.t.\ & 1 \leq x(i,O) \leq 2,\ i = i_1,\ i_2, \\ & x(N\times O) \leq 3,\ \ 0 \leq x \leq 1. \end{aligned}$$

An optimal solution is given by $x^*(i_1, o_1) = x^*(i_2, o_1) = x^*(i_2, o_3) = 1$, and $x^*(i, j) = 0$ otherwise. Note that at x^* the floor constraint for i_1, the ceiling constraint for i_2, and the aggregate capacity constraint are all binding. Similarly, we calculate β for the other type profiles. The corresponding implementability condition is given by

$$Q_{1,2}^2 + Q_{2,1}^1 + Q_{2,1}^2 + Q_{2,1}^3 - Q_{1,1}^1 - Q_{1,1}^2 - Q_{1,1}^3 - Q_{2,2}^2 \leq \lambda_{11} - \lambda_{12} + 3\lambda_{21} + \lambda_{22},$$

where $Q_{i,k}^j := Q_i(t_i^k, o_j)\lambda_i(t_i^k)$ and $\lambda_{kl} := \lambda(t_1^k, t_2^l)$.

Now we will present our second major result (i.e., Theorem 2 below), which identifies two classes of constraint structures $\mathcal{G}$ under which the projection cone of the implementation problem preserves total unimodularity. If $\mathcal{G}$ is a hierarchy, then the implementation structure consists of two hierarchies and by a well-known theorem of Edmonds (1970), the projection cone can be shown to preserve total unimodularity. However, when $\mathcal{G}$ is a bihierarchy, the implementation structure has three hierarchies, and the Edmonds theorem cannot be applied anymore. Fortunately, we can still show that the projection cone preserves total unimodularity in this case if two additional conditions are imposed: (1) $\mathcal{G}$ contains also both $\{i\} \times A$ for every $i \in N$ and $N \times \{j\}$ for every $j \in A$; and (2) the set T_i for every $i \in N$ contains at most two elements, which is called binary.

Theorem 2. *Let $(N, A, \mathcal{G}, T)$ be an implementation structure. The projection cone preserves total unimodularity, if one of the following conditions holds:*

(1) $\mathcal{G}$ is a hierarchy.

(2) $\mathcal{G}$ is a bihierarchy and contains also $\{i\} \times A$ for every $i \in N$ and $N \times \{j\}$ for every $j \in A$, and the set T_i for every $i \in N$ has at most two elements.

Proof. We only need to prove (2). Without loss of any generality, we consider a case where each agent has exactly two types. Let C_i be the submatrix of C for the constraint sets in $\mathcal{G}_i$, $i = 1, 2$. Let B denote the constraint matrix of the reduced-form implementation equalities. Since total unimodularity is preserved by deleting unitary column and by transpose, we only need to show that the following matrix

$$M^* = \begin{bmatrix} C_1 \\ C_2 \\ B \end{bmatrix} \tag{11}$$

is totally unimodular. We first prove the result for the problem with the standard constraints, i.e., $\mathcal{G}$ consists exactly of $\{i\} \times A$ for every $i \in N$ and $N \times \{j\}$ for every $j \in A$ and every singleton $(i, j) \in N \times A$, and the set T_i for every $i \in N$ has exactly two elements. Then we will show that all other cases can be reduced to the cases with the standard constraints.

(a) Standard constraints. In this case, each column of M^* contains exactly three 1s (by leaving the rows of singletons out). Note that if a matrix with at most three nonzero entries in each column, then A is TUM if and only if each submatrix of A with at most two nonzero entries in each column is TUM (Schrijver, 1986, Truemper, 1985). To show that M^* is TUM, we only need to show that each submatrix M of M^* with at most two nonzero entries in each column is TUM. By a theorem of Camion (Schrijver, 1986), if M contains at most two nonzeros in each column, M is TUM if and only if M is balanced, i.e., in every square submatrix M' with exactly two nonzero entries per row and per column, the sum of the entries is a multiple of four. We need to show that M' must contain an even number of rows.

Note each column and each row in M' contains two 1s. Connecting the nonzero entries in each row and column implies that M' can be decomposed into n disjoint cycles, where each cycle contains a submatrix of M' with exactly two 1s entries per row and per column. We show that each cycle P is even order.

First note that for any pair of two 1s in each row of P, there are three types of possible changes of column indexes: (i) from (i, j, t) to (i, j', t) in row (i, t) of C_1, and (ii) from (i, j, t) to (i', j, t) in row (j, t) of C_2, and (iii) from (i, j, t_i, t_{-i}) to (i, j, t_i, t'_{-i}) in row (i, j, t_i) of B. We can classify a cycle P into one of the following two cases:

Case 1. All column index changes in P belong to only two types of (i)-(iii). Then we must have the index changes are alternating between the two types. It implies that there is an equal number of index changes for the two types. That is, P must be a cycle of even order.

Case 2. All column index changes in P belong to all three types of (i)-(iii). First note the number of index changes of type (iii) (i.e., $t_{-i} \to t'_{-i} \to ... \to t_{-i}$) in the cycle is even since each player has two types. Furthermore, the number of index changes of type (i) and (ii) in total (i.e., $(i, j) \to (i', j) \to (i', j') \to ... \to (i, j)$) is also even as there are two hierarchies. So the total number of index changes of all three types is even. Hence P is a cycle of even order.

Combining the two cases, we conclude that each cycle P contains an even number of rows. This completes the proof that M' is balanced, and M (and M^*) is TUM.

(b) General constraints. Pick any square submatrix M of M^*. It is sufficient to prove that $det(M) \in \{0, \pm 1\}$. We claim that after some elementary row operations M can be converted into a matrix $\tilde{M}$ where for each column, there is at most one 1 for each hierarchy. Specifically, for any two rows of C_l ($l = 1, 2$), (k, t) and (k', t), let G_k and $G_{k'}$ be the corresponding sets in $\mathcal{G}_l$. Suppose $G_k \subset G_{k'}$. Negate row (k, t) and add it to row (k', t). The elementary row operation changes only the sign of the determinant of M and hence $|det(M)| = |det(\tilde{M})|$. It is sufficient to show that $\tilde{M}$ is TUM. Since $\tilde{M}$ now contains at most three 1s in each column, we only need to show that each submatrix of $\tilde{M}$ with at most two nonzero entries in each column is TUM. Following the proof of part (a), we obtain this result. □

We next reduce the condition in Theorem 1 to more specific ones. First note that when there are only floor or ceiling constraints, W is pointed: $(f, g, h)M(W) = 0$ implies $(f, g, h) = 0$. Hence W is generated by the set of extreme rays. We derive an explicit characterization of the set $\hat{W}$ of $\{0, \pm 1\}$ extreme rays. In particular, for each of the following cases, we obtain a characterization of $\hat{W}$: (1) There are only floor constraints, i.e., $c = +\infty$ (Theorem 3); (2) There are only ceiling constraints, i.e., $c < +\infty$ and $b = 0$ (Theorem 4).

To state the characterization results, we introduce some additional definitions. For every $G \subseteq N \times A$, define an effective lower bound $\bar{b}(G) := -\beta(\emptyset, G)$ and an effective upper bound $\bar{c}(G) := \beta(G, \emptyset)$. We also define $\Gamma(S) = \{t \in T \mid (t_i, j) \in S_i, \text{ for some } i \in N\}$. We first present a characterization result concerning floor constraints.

Theorem 3. (Floor constraints) *Suppose the projection cone preserves total unimodularity. Then Q is implementable if and only if for all i and $(t_i, j) \in T_i \times A$, $Q_i(t_i, j) \geq 0$, and for all $S_i \subseteq T_i \times A$, $i \in N$,*

$$\sum_{i \in N} \sum_{(t_i, j) \in S_i} Q_i(t_i, j)\lambda_i(t_i) \geq \sum_{t \in \Gamma(S)} \lambda(t)\bar{b}(I(t, S)). \tag{12}$$

The next result gives a characterization for the case with ceiling constraints.

Theorem 4. (Ceiling constraints) *Suppose the projection cone preserves total unimodularity. Then Q is implementable if and only if for all i and $(t_i, j) \in T_i \times A$, $Q_i(t_i, j) \geq 0$, and for all $S_i \subseteq T_i \times A$, $i \in N$,*

$$\sum_{i \in N} \sum_{(t_i,j) \in S_i} Q_i(t_i, j)\lambda_i(t_i) \leq \sum_{t \in \Gamma(S)} \lambda(t)\bar{c}(I(t, S)). \tag{13}$$

Several remarks are in order. First, when $\mathcal{G}$ is a hierarchy, condition (13) reduces to the implementability condition obtained by Che et al. (2013) who study the case of multiple units of one good. The condition further reduces to the classical condition of Border (1991) for a single object, i.e., $\mathcal{G} = \{N\}$ and $c(N) = 1$. Second, although the implementability conditions of Theorems 3 and 4 appear to be similar, the projection cones of the two problems and their extreme rays have very different structures (See Lemmas 5 and 6). Hence the two problems have to be treated separately and also they concern quite different constraints. Third, the conditions in Theorems 3 and 4 characterize the implementability for a general class of implementation structures. As we have shown in Theorem 2, our results allow for hierarchical families and some class of bihierarchical families.

4 The Lift-and-Project Approach

In this section we present the mathematical method, i.e., the lift-and-project approach, to establish our characterization results introduced in the previous section. This is a powerful method in polyhedral combinatorics (Balas and Pulleyblank, 1983) and has been explored by Vohra (2013) in economics. The basic idea of the method goes as follows: The first step is to use a linear system in some higher-dimensional space, or lifting. The second step is to obtain a linear system by properly projecting away previously added variables. We have tried to keep the presentation as simple as possible while maintaining rigor.

First we describe a general lift-and-project method. Suppose we are given a polyhedron

$$Z = \{(x, y) \in \mathbb{R}^p \times \mathbb{R}^q \mid A^1x + B^1y = b^1,\ A^2x + B^2y \leq b^2\}, \tag{14}$$

where A^1, B^1, A^2, B^2 are matrices, b^1, b^2 are vectors. Let Y denote the projection of Z onto the subspace of y variables, that is,

$$Y = \{y \in \mathbb{R}^q \mid \text{ there exists } x \in \mathbb{R}^p \text{ such that } (x, y) \in Z\}. \tag{15}$$

We wish to obtain a linear system whose solution set is Y.

We define the projection cone

$$P = \{(f, g) \in \mathbb{R}^r \times \mathbb{R}^s \mid f^\top A^1 + g^\top A^2 = 0, g \geq 0\}. \tag{16}$$

Let $\hat{P}$ be any finite set of generators of P. The following lemma shows that finding the linear inequalities that define Y reduces to finding finite generators of P.

Lemma 1. *The projection of the polyhedron Z given by* (14) *onto y is given by*

$$Y = \{y \in \mathbb{R}^q \mid (f^\top B^1 + g^\top B^2)y \leq f^\top b^1 + g^\top b^2, \textit{for all } (f, g) \in \hat{P}\}.$$

Characterizing the generators of P appears to be difficult in general. The following lemma provides a sufficient condition on P such that a complete description of the generators can be found. The lemma is developed from and slightly more general than a result in Hoffman (1976), where the result there is proved for pointed cones. We show that the same result holds also for non-pointed cones.

Lemma 2. *If the constraint matrix of the projection cone P given by* (16) *is totally unimodular, then P is generated by a set $\hat{P}$ of $\{0, \pm 1\}$ generators.*

Remark 1. If P is pointed, the set of extreme rays provides a unique (up to positive scaling) minimal set of generators. While Lemma 1 shows that the extreme rays of P provide a complete description of Y, some extreme ray may be redundant.

Remark 2. Lemma 2 is useful on its own since it identifies finitely many linear inequalities. In particular, it implies that we can pick $\hat{P}$ to be all $\{0, \pm 1\}$ vectors in P, i.e., $P \cap \{0, \pm 1\}^{r+s}$.

Remark 3. While total unimodularity is sufficient for $\{0, \pm 1\}$ generators, it is not necessary. For example, let P be the cone defined by $x_1 + x_2 \leq 0$ and $x_2 \leq x_1$. The constraint matrix of P is not totally unimodular, but its extreme rays are $\{0, \pm 1\}$ vectors.

To prove Theorem 1, the first step is to apply the lift-and-project method to our implementation problem. Note that if the projection cone is totally unimodular, the implementation system can be described by the linear constraints F in (6). Then Lemma 1 applies to the implementation system F. Since the projection cone W in (7) is totally unimodular, applying Lemma 2 to W we obtain Lemma 3 below, which gives a characterization for implementation.

Lemma 3. *Let $(N, A, T, \lambda, \mathcal{G}, b, c)$ be an implementation problem. Suppose the projection cone preserves total unimodularity. Then $y \in \mathbb{R}^l$ is implementable if and only if*

$$f^\top y \leq g^\top c^\lambda - h^\top b^\lambda, \text{ for all } (f, g, h) \in \hat{W}, \tag{17}$$

where the set $\hat{W}$ consists of generators of W given by (7) with entries $-1, 0,$ and $+1$.

The following lemma shows how the bound in the right hand side of condition (17) reduces to the effective bound β defined in (9), which is determined by the dual of a linear programming problem. This allows us to obtain explicit bounds in Theorem 1. Using these lemmas together will establish Theorem 1 whose proof is deferred to the appendix.

Lemma 4. *For any pair of disjoint sets $G, H \subseteq N \times A$, the effective bound $\beta(G, H)$ defined in (9) is equal to the minimum value of a linear programming problem. That is,*

$$\begin{aligned}
\beta(G, H) = \min_{g,h} & \sum_{U \in \mathcal{G}} c(U)g(U) - b(U)h(U) \\
s.t. & \sum_{U \in \mathcal{G}} g(U)\chi^U - \sum_{U \in \mathcal{G}} h(U)\chi^U = sign^{G,H}, && (18) \\
& g, h \geq 0, && (19)
\end{aligned}$$

where $\chi^U(i, j) = 1$ if $(i, j) \in U$ and 0 otherwise; $sign^{G,H}(i, j) = 1$ if $(i, j) \in G$, -1 if $(i, j) \in H$, and 0 otherwise.

Theorems 3 and 4 will be shown in the Appendix by using a series of lemmas below. We first analyze the problem with floor constraints. By setting $g = 0$ in Lemma 3, the projection cone is given by

$$W = \{(f, h) \in \mathbb{R}^{l+r} \mid f^\top B + h^\top C = 0, h \geq 0\}. \tag{20}$$

Observe that W is a pointed cone. The next lemma provides a characterization of the set $\hat{W}$ of $\{0, \pm 1\}$ extreme rays of W. For any $U \subseteq N \times A$, a *cover* $\mathcal{P}$ of U is defined by a collection of constraint sets in $\mathcal{G}$ such that $U \subseteq \cup_{G \in \mathcal{P}} G$. We say a cover $\mathcal{P}$ is *minimal*, if there exists no other cover $\mathcal{P}'$ such that $\mathcal{P}' \subset \mathcal{P}$.

Lemma 5. (Floor constraints) (f, h) *is a* $\{0, \pm 1\}$ *extreme ray of the set* W *given by* (20) *if and only if* $(-f, h)$ *is the incidence vector of some* $(S, (\mathcal{P}_t)_{t \in T})$*, where* $S \subseteq \Omega$ *and each* $\mathcal{P}_t$ *is a minimal cover of* $I(t, S)$ *with* $\cup_{G \in \mathcal{P}_t} G = I(t, S)$*. That is,*

$$f(i, t_i, j) = \begin{cases} -1, & \text{if } (t_i, j) \in S_i, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$h(G, t) = \begin{cases} +1, & \text{if } t \in \Gamma(S) \text{ and } G \in \mathcal{P}_t, \\ 0, & \text{otherwise.} \end{cases}$$

We now turn to the problem with only ceiling constraints. By Lemma 3, the value of h does not affect the implementability inequality. By eliminating $h \geq 0$ in W, the projection cone can be written as

$$W = \{(f, g) \in \mathbb{R}^{l+r} \mid -f^\top B + g^\top C \geq 0, g \geq 0\}. \tag{21}$$

Note that W is pointed and we obtain the following characterization of the set $\hat{W}$ of $\{0, \pm 1\}$ extreme rays.

Lemma 6. (Ceiling constraints) (f, g) *is a* $\{0, \pm 1\}$ *extreme ray of the set* W *given by* (21) *if and only if either of the following conditions holds:*

(1) $f(i, t_i, j) = -1$ *for exactly one* (i, t_i, j) *and 0 otherwise, and* $g = 0$.

(2) (f, g) *is the incidence vector of some* $(S, (\mathcal{P}_t)_{t \in T})$*, where* $S \subseteq \Omega$ *and each* $\mathcal{P}_t$ *is a minimal cover of* $I(t, S)$*, that is,*

$$f(i, t_i, j) = \begin{cases} +1, & \text{if } (t_i, j) \in S_i, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$g(G, t) = \begin{cases} +1, & \text{if } t \in \Gamma(S) \text{ and } G \in \mathcal{P}_t, \\ 0, & \text{otherwise.} \end{cases}$$

In each of Lemmas 5 and 6, it follows that the set of extreme rays $\hat{W}$ is characterized by the incidence vectors of the form $(S, (\mathcal{P}_t)_{t\in T})$. On the other hand, some extreme rays are redundant in describing the implementability condition. In the Appendix, we complete the proof of Theorems 3 and 4 by removing redundant extreme rays.

5 Case studies

In this section we discuss several interesting cases where our approach and results can be applied. We will briefly examine each case.

5.1 Hierarchical structures

In this subsection, we investigate the structures of feasible reduced forms and obtain characterizations that are useful for optimization and implementation. Vohra (2011) shows that the set of feasible reduced form auctions is a polymatroid, which implies that feasible reduced forms that optimize over a given linear objective can be found by the greedy algorithm. Alaei et al. (2019) develop a polymatroidal decomposition property to show that the set of feasible reduced forms is a polymatroid associated with an expected rank function.[2] We intend to investigate the polyhedral aspect of the implementability condition in our environment. For illustration, we restrict attention to hierarchies with ceiling constraints only.

Let Ω be the ground set. For every $S \subseteq \Omega$ and $t \in T$, define $f^t(S) = \bar{c}(I(t, S))$. For each $S \subseteq \Omega$, define

$$f(S) = \sum_{t\in T} \lambda(t) f^t(S).$$

We call f^t an ex post rank function and f the expected rank function. By the change of variables, we denote $y(t_i, j) = Q_i(t_i, j)\lambda_i(t_i)$. By Theorem 4, the set of implementable reduced forms in (13) can be written as

$$\mathcal{Q} = \{y \in \mathbb{R}^{\Omega}_{+} \mid \sum_{e\in S} y(e) \le f(S), \text{ for all } S \subseteq \Omega\}. \tag{22}$$

Proposition 1. *If $\mathcal{G}$ is a hierarchy, the set $\mathcal{Q}$ given by (22) is a polymatroid.*

[2] They also provided computationally tractable methods for optimization and implementation of interim allocation rules.

We now characterize the extreme points of the feasible reduced forms. For symmetric reduced-form auctions, Border (1991) constructs a class of hierarchical allocation rules that implement the extreme points of feasible reduced forms. Alaei et al. (2019) introduce a class of ordered subset allocation rules for implementation in auction problems with matroid constraints. One common feature of hierarchical and ordered subset allocations is that both use some priorities to rank agents by their types. We generalize the ordered subset allocations of Alaei et al. (2019) to our environment.

Definition 4. *A subset $\Pi \subseteq \Omega$ is an ordered set if Π is given by an ordering on elements $((t^1, j^1), (t^2, j^2), ..., (t^l, j^l))$, where (t^k, j^k) denotes the k-th element in Π.*

Note that priorities depend not only on types but also on objects, i.e., while type t_i of agent i may have a higher priority for object j than the other agents' types in the list, it may have a lower priority for some other object j'.

Definition 5. (Ordered subset mechanisms) *A feasible mechanism p is an ordered subset mechanism for an ordered subset Π of Ω, if the mechanism orders the agents on the basis of their type-object pairs according to Π and allocates the objects sequentially, given the feasibility constraints defined by f^t. That is,*

$$p(i, j, t) = \begin{cases} 1, & \textit{if } (t_i, j) \in S^k \setminus S^{k-1} \textit{ and } f^t(S^k) \geq f^t(S^{k-1}) + 1, \\ 0, & \textit{otherwise.} \end{cases}$$

where $S^k = \{(t^1, j^1), ..., (t^k, j^k)\}$ denotes the first $k \leq |\Pi|$ elements of Π.

For the ordered subset mechanism we have the following characterization result.

Proposition 2. *Every extreme point of the set $\mathcal{Q}$ given by* (22) *is implementable by an ordered subset mechanism.*

We will use Example 1 to illustrate an ordered subset mechanism.

Example 8. (Example 1 continued) Recall that there are three doctors $\{j_1, j_2, j_3\}$ and three hospitals (agents) $N = \{i_1, i_2, i_3\}$, and the regional cap requires at most two doctors being allocated. The effective capacity is given by $\bar{c}(G) = 1$ if $G \subseteq N \times \{j\}$ for some j and $\bar{c}(G) = 2$ otherwise. Let

$$\Pi = ((t_1^1, j_1), (t_2^1, j_1), (t_3^1, j_2), (t_3^2, j_1), (t_1^1, j_3))$$

be an ordered subset.

For this example, with respect to the ordered subset mechanism, at type profile $t = (t_1^1, t_2^1, t_3^1)$, we have $I(t, \Pi) = \{(i_1, j_1), (i_2, j_1), (i_3, j_2), (i_1, j_3)\}$ and $p(i, j, t)$ is given as follows: At round 1, agent i_1 gets j_1; At round 2, agent i_2 gets nothing as $\{(i_1, j_1), (i_2, j_1)\}$ is infeasible; At round 3, agent i_3 gets j_2; and at round 4, agent i_1 gets nothing as $\{(i_1, j_1), (i_3, j_2), (i_1, j_3)\}$ is not feasible.

5.2 Bilateral Trade

We show how the conic method can be applied to the classic bilateral trade problem of Myerson and Satterthwaite (1983). Observe that the bilateral trade problem corresponds to the case of our current model where there are two players and one unit of a single good, and where the ceiling and the floor are equal. In particular, we set $N = \{1, 2\}$, $|A| = 1$, $\mathcal{G} = \{N\}$, $b(N) = c(N) = 1$, and $b(\{i\}) = 0$. Note that the equality constraint implies that g is free in the set W given by (7). By eliminating $h \geq 0$, the projection cone reduces to

$$W = \{(f, g) \in \mathbb{R}^{l+r} \mid -f^\top B + g^\top C \geq 0\}.$$

We characterize the generators of the projection cone for this bilateral trade problem. In contrast to the auction problem whose projection cone is *pointed*, we show that the projection cone of a bilateral trade is *non-pointed.*

Lemma 7. *For bilateral trade, we have: (1) W is non-pointed. (2) $M(W)$ is TUM.*

The next result gives a detailed characterization of the projection cone concerning bilateral trade.

Lemma 8. *For bilateral trade, if (f, g) is a $\{0, \pm 1\}$ generator of W, then one of the following conditions holds:*

(1) *$(f, g) = (1, ..., 1)$ or $(f, g) = (-1, ..., -1)$.*
(2) *$f(i, t_i, j) = -1$ for a unique (i, t_i, j) and 0 otherwise, and $g = 0$.*
(3) *(f, g) is the incidence vector of some $(E, \Gamma(E))$, where $E_i \subseteq T_i$ for all $i \in N$.*

That is,

$$f(i, t_i) = \begin{cases} +1, & \textit{if } t_i \in E_i, \\ 0, & \textit{otherwise,} \end{cases}$$

and

$$g(t) = \begin{cases} +1, & \textit{if } t \in \Gamma(E), \\ 0, & \textit{otherwise.} \end{cases}$$

The generators in condition (1) of the lemma correspond to a basis of the linearity space of the projection cone. These generators correspond to the equality constraint in the implementability condition:

$$\sum_{i=1,2} \sum_{t_i \in T_i} Q_i^i(t_i)\lambda_i(t_i) = \sum_{t \in T} \lambda(t). \tag{23}$$

Note that if the projection cone is non-pointed, the linear description is not unique and (23) can be used to generate different descriptions. For bilateral trade, as both the seller and the buyer are interested in the interim expected probability of trade (i.e., $Q_i^1(t_i)$), we can obtain the following characterization result.

Proposition 3. *For bilateral trade, $Q \in \mathbb{R}_+^l$ given by* (23) *is implementable if and only if*

$$\sum_{t_1 \in T_1} Q_1^1(t_1)\lambda_1(t_1) - \sum_{t_2 \in T_2} Q_2^1(t_2)\lambda_2(t_2) = 0, \tag{24}$$

and for all E with $E_i \subseteq T_i$ for all $i = 1, 2$,

$$\sum_{t_1 \in E_1} Q_1^1(t_1)\lambda_1(t_1) - \sum_{t_2 \in E_2} Q_2^1(t_2)\lambda_2(t_2) \leq \lambda(E_1 \times E_2^c). \tag{25}$$

That is, at the ex ante stage, the two players must have the same expectations on the probability of trade, and the difference in the seller's and the buyer's interim probabilities of trade for any set of types $E_1 \times E_2$ cannot be too distinct.

We provide some interpretation of the implementability conditions, by comparing an auction with a bilateral trade. In Border's theorem, only coefficients "+1" appear in the linear inequalities. In contrast, the implementability condition for bilateral trade here contains coefficients not only "+1" but also "−1" in the linear inequalities.

The interpretation for this result is intuitive: In Myerson (1981), the "+1" coefficient means that if for buyer 1 the expected probability of winning becomes higher, then for buyer 2 the expected probability of winning must be lower as the buyers are competing for the probabilities of winning. In Myerson and Satterthwaite (1983), however, the "+1" coefficient refers to that for each player her expected probability of trade for some types is obtained by summing up her interim probabilities of trade for these types, the "−1" coefficient means that the difference between the seller's ex ante expected probability of trade and the buyer's cannot be too large. This is because trade is a public alternative: increasing the probability that trade occurs would increase both players' expected probabilities of trade.

5.3 Compromise

We discuss how to apply our approach to the compromise problem studied by Börgers and Postl (2009). In their model there are two players (i.e., $N = \{1, 2\}$) and three alternatives (i.e., $A = \{a_0, a_1, a_2\}$). The players have opposite preferences: $a_1 \succ_1 a_0 \succ_1 a_2$ and $a_2 \succ_2 a_0 \succ_2 a_1$. That is, the best alternative to one player is the worst alternative to the other player. We normalize $u_i(a_0) = 0$ for $i = 1, 2$. Each player i has private information about her payoffs on a_1 and a_2, given by a type $t_i \in T_i$.[3]

While the number of types in Börgers and Postl (2009) can be any positive integer, for illustrative purpose we focus on the case of two types, i.e., $|T_1| = |T_2| = 2$. In this problem a feasible allocation rule $q : T \to \Delta(A)$ assigns each type profile a lottery over alternatives. Hence for each player with type t_i, the reduced form allocation probability is multidimensional: $Q_i(t_i) = (Q_i(t_i, a_1), Q_i(t_i, a_2))$.

To apply the conic method in this setting, let B be an incidence matrix where each row is indexed by (i, a, t_i), each column is indexed by (a, t), and the entry in row (i, a, t_i) and column (a', t') is 1 if $a = a'$ and $t_i = t'_i$ and 0 otherwise. Let C be an incidence matrix where each row is indexed by t, each column is indexed by (a, t), and the entry in row t and column (a', t') is 1 if $t = t'$ and 0 otherwise. The

[3]Note that Börgers and Postl (2009) normalize $u_i(a_i) = 1$ and $u_i(a_j) = 0$ and assume each player has private information about her payoff on the compromise alternative k_0, i.e., $u_i(a_0) = t_i$. The reduced-form implementation problem is the same irrespective of the normalizations.

projection cone is given by

$$W = \{(f, g) \mid -f^{\top} B + g^{\top} C \geq 0\}.$$

We will prove that when each player has two types, the projection cone is non-pointed and the constraint matrix of the projection cone is totally unimodular.

Lemma 9. *For compromise with binary type sets, we have: (1) W is non-pointed. (2) $M(W)$ is TUM.*

It is worth noting that while $M(W)$ does not form two laminars (it contains three laminars), which differs from the bilateral trade problem, the constraint matrix remains totally unimodular.

The above result implies that the projection cone is generated by $\{0, \pm 1\}$ vectors. To present the implementability condition, let $T_1 = \{i_1, i_2\}$, $T_2 = \{j_1, j_2\}$, and $K = \{a_1, a_2\}$. Denote $Q_{il}^a = Q_i(l, a)\lambda_i(l)$. It can be verified that the implementability inequalities are given by

$$\frac{1}{2}\sum_{a \in K}(Q_{1i}^a - Q_{1i'}^a + Q_{2j_1}^a + Q_{2j_2}^a) \leq \lambda_{ij_1} + \lambda_{ij_2}, \quad \forall i \in T_1, \tag{26}$$

$$\frac{1}{2}\sum_{a \in K}(Q_{1i_1}^a + Q_{1i_2}^a + Q_{2j'}^a - Q_{2j}^a) \leq \lambda_{i_1j} + \lambda_{i_2j}, \quad \forall j \in T_2, \tag{27}$$

$$\frac{1}{2}\sum_{a \in K}(Q_{1i}^a - Q_{1i'}^a + Q_{2j'}^a - Q_{2j}^a) \leq \lambda_{ij}, \quad \forall (i, j) \in T, \tag{28}$$

$$\frac{1}{2}(Q_{1i}^a - Q_{1i'}^a + Q_{2j'}^a - Q_{2j}^a) \leq \lambda_{ij}, \quad \forall (a, i, j) \in K \times T. \tag{29}$$

6 Concluding Remarks

Many practical problems and markets face various complex distributional constraints going beyond the traditional ceiling or capacity constraints. In this paper, we have studied the implementation problem of reduced-form allocation of multiple indivisible objects to many agents with distributional constraints. In our model, objects can be private goods or public/club goods, every agent may demand several objects and has private information over her preferences. Her private information is described by a finite set of types. Distributional constraints are described by a variety of families of pairs of agent and object.

Using a lift-and-project method, we have been able to obtain a conic approach for studying the implementability condition. We have shown how this conic approach can be applied to many important problems including matching, auctions, bilateral trade, compromise with distributional constraints. We have demonstrated how the approach allows for a unified treatment of different classes of problems. We have succeeded in identifying a fundamental condition called total unimodularity and establishing a general characterization result on implementation under this condition. Total unimodularity reflects the essential property of the class of well-studied totally unimodular matrices. Analyzing these matrices offers also interesting criteria that can be explored to classify different classes of economic problems. For each problem, the main task is to check whether its projection cone preserves total unimodularity or not. In fact, we have proved that two large classes of constraint structures: hierarchies and bihierarchies with canonical constraints, can ensure total unimodularity.

We hope this study has shed new light on implementation of reduced-form allocation and will provide a useful and necessary basis for further study on many complex real life resource allocation problems.

Appendix

Proof of Theorem 1. We prove this result with the following steps. First note that if the implementation structure is TUM, then the constraint matrix of $\mathcal{G}$ is TUM, as every submatrix of a TUM matrix is TUM. Hence, the set of feasible ex post random allocations has a linear characterization defined by condition (1), and the linear constraints F describe the implementation system. Next, since the projection cone is TUM, we apply Lemma 2 to F and know that the generators of the projection cone are $\{0, \pm 1\}$ vectors. By Lemma 3, pick any non-zero $f \in \{0, \pm 1\}^l$ such that $(f, g, h) \in W$ for some $g, h \geq 0$. f corresponds to the incidence vector of some $(S_i^+, S_i^-)_{i \in N}$. Denote $W(f) = \{(g, h) : (f, g, h) \in W\}$. Consider the problem of minimizing $g^\top c^\lambda - h^\top b^\lambda$ subject to $(g, h) \in W(f)$. Note that the problem reduces to the following pointwise optimization, i.e., for every $t \in T$,

$$\min g_t^\top c - h_t^\top b, \ s.t. \ (g_t, h_t) \in W_t(f), \tag{30}$$

where $g_t, h_t, W_t(f)$ are the projections of $g, h, W(f)$ onto t. That is, $W_t(f)$ is given by

$$\sum_{G\in\mathcal{G}} g_t(G)\chi^G - \sum_{G\in\mathcal{G}} h_t(G)\chi^G = sign^{I(t,S^+),I(t,S^-)},$$

$$g_t(G), h_t(G) \geq 0,$$

where $sign^{I(t,S^+),I(t,S^-)}(i,j) = 1$ if $(i,j) \in I(t,S^+)$, -1 if $(i,j) \in I(t,S^-)$ and 0 otherwise. By Lemma 4, we have the value of the minimization problem is equal to $\beta(I(t,S^+), I(t,S^-))$. □

Proof of Lemma 1. We first deal with the 'only if' part. Suppose y is implementable by some x. Pick any $(f,g) \in P$ and $f^\top A^1 + g^\top A^2 = 0$, $g \geq 0$. Then $(f^\top B^1 + g^\top B^2)y \leq -(f^\top A^1 + g^\top A^2)x + f^\top b^1 + g^\top b^2 = f^\top b^1 + g^\top b^2$. Hence $y \in Y$.

Now we turn to the 'if' part. Suppose y is not implementable. There exists no x such that $A^1x = b^1 - B^1y$ and $A^2x \leq b^2 - B^2y$. By Farkas' Lemma (Schrijver, 1986, p.89), there exists (f,g) such that $f^\top A^1 + g^\top A^2 = 0$, $g \geq 0$, and $f^\top(b^1 - B^1y) + g^\top(b^2 - B^2y) < 0$. But then $(f,g) \in P$. There must be $(\hat{f}, \hat{g}) \in \hat{P}$ such that $(\hat{f}^\top B^1 + \hat{g}^\top B^2)y > \hat{f}^\top b^1 + \hat{g}^\top b^2$, and hence $y \notin Y$. □

Lemma 2 is immediately obtained by using the following mathematical result, which gives a nice and clear characterization of general cones defined by a totally unimodular matrix (TUM) and will be derived from a well-known lemma of Hoffman (1976).

We first introduce some notations. Let I be the index set of the inequalities in the cone P. Let $J = I^=(x)$ denote the index set in I for which the corresponding inequalities hold as equations (or active constraints) at $x \in P$. Let M_J be the corresponding submatrix for J, and let $\text{rk}M_J$ denote the rank of M_J. Note that $x \in P$ is an extreme ray if and only if $I^=(x)$ is maximal, i.e., there exists no $x' \in P$ such that $J' = I^=(x')$ and $J \subset J' \subset I$. If x is an extreme ray, then so is λx for all $\lambda \geq 0$. Observe that for a pointed cone, the set of extreme rays provides a unique (up to positive scaling) minimal set of generators.

Lemma 10. *Let $P = \{x \in \mathbb{R}^p \mid Mx \leq 0\}$ be a polyhedral cone and let M be TUM.*

(1) If P is pointed, then P is generated by $\{0, \pm 1\}$ extreme rays.

(2) If P is non-pointed, then P is generated by $\{0, \pm 1\}$ vectors.

Proof of Lemma 10. (1) Let P be pointed. Assume that $z \in P$ is an extreme ray and $J = I^{=}(z)$. Since $\mathrm{rk} M_J = q - 1$, there is a submatrix M' with $q-1$ linearly independent rows in M_J. Since M is TUM, M' is also TUM. We need to show that if M' with columns $M'_1, ..., M'_q$ has rank $q-1$, then $M'z = 0$ implies that all nonzero coordinates of z are either α or $-\alpha$, for some $\alpha > 0$. The proof is essentially the same as Lemma 3.1 of Hoffman (1976). Note that for any $z_j \neq 0$,

$$M'_j = \sum_{i=1, i \neq j}^{q} \frac{z_i}{z_j} M'_i, \tag{31}$$

and M'_i, $i \neq j$ are linearly independent. The linear system has a unique solution $(\frac{z_i}{z_j})$, $i \neq j$. Since M' is TUM, each $\frac{z_i}{z_j}$ is integer. As this argument applies to every nonzero entry in z, for any nonzero $z_j, z_{j'}$ and $j \neq j'$, $\frac{z_{j'}}{z_j}$ and $\frac{z_j}{z_{j'}}$ are integers and hence $|z_j| = |z_{j'}| = \alpha$ for some $\alpha > 0$. Therefore every extreme ray of P contains a $\{0, \pm 1\}$ vector.

(2) Let P be non-pointed. We write P as a union of finitely many pointed cones. Notice that the Euclidean space $\mathbb{R}^q$ is a union of $l := 2^q$ closed orthants $i = 1, ..., l$. Let P_i be the intersection of P and orthant i (with some P_i possibly empty). Then $P = \cup_{i=1}^{l} P_i$. We claim that each P_i and hence P is generated by $\{0, \pm 1\}$ vectors. By the first part of the lemma, we only need to show that for each i, P_i is a pointed cone and its constraint matrix M_i is TUM. Notice that by construction, each P_i is a pointed cone. Since M is TUM and M_i is obtained from M by adding rows with at most one non-zero entries (1 or -1), M_i is also TUM. □

Proof of Lemma 3. Applying Lemmas 1 and 2 to F and W yields immediately the result. □

Proof of Lemma 4. $\beta(G, H)$ is defined by the following problem

$$\beta(G, H) = \max\{w^\top x \mid b \leq x \leq c\}, \tag{32}$$

where $w = sign^{G,H}$. The dual problem of the problem above is given by

$$\min g^\top c - h^\top b \ s.t. \ (18) \text{ and } (19). \tag{33}$$

By strong duality, the existence of an optimal solution for the primal problem implies that the dual problem also has an optimal solution and the optima are equal. $\square$

We need to introduce the following result for the proof of Theorem 3.

Lemma 11. *Suppose there are floor constraints only. For each $U \subseteq N \times A$,*

$$\bar{b}(U) = \max\{\sum_{G\in\mathcal{P}} b(G) \mid \mathcal{P} \subseteq \mathcal{G} \text{ is a minimal cover of } U \text{ with } \cup_{G\in\mathcal{P}} G = U\}. \quad (34)$$

Proof. For any $U \subseteq N \times A$, $\bar{b}(U)$ is defined as the value of the following problem

$$\min\{x(U) \mid x(G) \geq b(G), \forall G \in \mathcal{G}\}. \quad (35)$$

The dual problem of the problem above is given by

$$\max\{\sum_{G\in\mathcal{G}} b(G)h(G) \mid h \geq 0, \sum_{G\in\mathcal{G}} h(G)\chi^G = \chi^U\}. \quad (36)$$

By strong duality, the primal has a solution implies the dual problem also has a solution and the optima are equal. Without loss, let $0 \leq h \leq 1$ be an optimal solution. Since the constraint matrix is totally unimodular, h can take 0-1 values. Hence h is the incidence vector of $\mathcal{P} \subseteq \mathcal{G}$. Moreover, $\mathcal{P}$ is a minimal cover of U with $\cup_{G\in\mathcal{P}} G = U$. $\square$

Proof of Lemma 5. We deal with the 'only if' part. Suppose $z = (f, h)$ is a $\{0, \pm 1\}$ extreme ray. $h \geq 0$ implies $f \leq 0$. Then $-f$ is the incidence vector of some S. That is, for each $t \in T$, $f(i, t_i, j) = -1$ for all $(i, j) \in I(t, S)$ and $f(i, t_i, j) = 0$ otherwise. Denote $h_t = h(G, t)_{G\in\mathcal{G}}$. Then $z \in W$ implies

$$\sum_{G\in\mathcal{G}} h_t(G)\chi^G = \chi^{I(t,S)}, \quad (37)$$

$$h_t(G) \geq 0. \quad (38)$$

Case 1: $t \notin \Gamma(S)$. If the set of active constraints is maximal, $h_t = 0$.

Case 2: $t \in \Gamma(S)$. Since each $h_t(G) \in \{0, 1\}$, h_t is the incidence vector of some $\mathcal{P}_t \subseteq \mathcal{G}$: $h_t(G) = 1$ for each $G \in \mathcal{P}_t$ and $h_t(G) = 0$ otherwise. (42) implies that each

(i, j) in $I(t, S)$ is contained in at least one $G \in \mathcal{P}_t$. Hence $\mathcal{P}_t$ is a cover. Moreover, each (i, j) in $I(t, S)$ is contained in exactly one $G \in \mathcal{P}_t$. Then $\mathcal{P}_t$ is minimal.

We turn to the 'if' part. Suppose z is not a $\{0, \pm 1\}$ extreme ray, i.e., the set of active constraints in (37)-(38) is not maximal at z. But then there exists z' such that more constraints in (38) are active. Since $\mathcal{P}'_t \subset \mathcal{P}_t$ is a cover of $I(t, S)$, we have $\mathcal{P}_t$ is not a minimal cover. □

Proof of Theorem 3. By Lemmas 5 and 11, for the extreme rays with given S, it implies that the extreme rays that does not yield the bounds $\bar{b}(I(t, S))$ for all $t \in T$ are redundant and we are done. □

The following result is used for the proof of Lemma 6 and Theorem 4.

Lemma 12. *Suppose there are ceiling constraints. For each $U \subseteq N \times A$,*

$$\bar{c}(U) = \min\{\sum_{G \in \mathcal{P}} c(G) \mid \mathcal{P} \subseteq \mathcal{G} \textit{ is a minimal cover of } U\}. \tag{39}$$

Proof. For any $U \subseteq N \times A$, $\bar{c}(U)$ is defined as the value of the following problem

$$\max\{x(U) \mid x \geq 0, x(G) \leq c(G), \forall G \in \mathcal{G}\}. \tag{40}$$

The dual problem of the problem above is given by

$$\min\{\sum_{G \in \mathcal{G}} c(G) g(G) \mid g \geq 0, \sum_{G \in \mathcal{G}} g(G) \chi^G \geq \chi^U\}. \tag{41}$$

A similar argument as Lemma 11 implies that g is the incidence vector of a minimal cover $\mathcal{P}$ of U. □

Proof of Lemma 6. We first deal with the 'only if' part. The 'if' part can be shown analogously as Lemma 5. Suppose $z = (f, g)$ is a $\{0, \pm 1\}$ extreme ray. There are two cases:

(1) f has at least one entry (i, t_i, j) with value -1. We show that all other coordinates of z must be 0. Note that $g \geq 0$ implies $f(i, t_i, j) + \sum_{G:(i,j) \in G} g(G, t) > 0$. Then in $M^* := M(W)$, each of columns (i, j, t) is not in $J = I^=(z)$. Let $\tilde{z} = (\tilde{f}, g)$ with $\tilde{f}(i, t_i, j) = 0$ and $\tilde{f} = f$ otherwise. Then $\tilde{z} M^*_J = 0$, i.e., for columns J the constraints remains active, but for columns (i, j, t) more constraints may become

active. Hence $J \subseteq \tilde{J} = I^{=}(\tilde{z})$. Since J is maximal, either $\tilde{J} = J$ or $\tilde{J} = I$. In the first case, $\tilde{z} = \alpha z$ for some $\alpha > 0$, yielding a contradiction. Hence, $\tilde{J} = I$ and $\tilde{z}M^* = 0$. Since W is pointed, $\tilde{z} = 0$ and $z = (0, ..., 0, -1, 0, ..., 0)$.

(2) $f \geq 0$. Then f is the incidence vector of some S: $f(i, t_i, j) = 1$ for all $(t_i, j) \in S_i$ and $f(i, t_i, j) = 0$ otherwise. That is, for each $t \in T$, $f(i, t_i, j) = 1$ for all $(i, j) \in I(t, S)$ and $f(i, t_i, j) = 0$ otherwise. We show that g must be the form in the Lemma. Fix $t \in T$ and denote $g_t = g(G, t)_{G \in \mathcal{G}}$. Then $(f, g) \in W$ implies

$$\sum_{G \in \mathcal{G}} g_t(G) \chi^G \geq \chi^{I(t,S)}, \tag{42}$$

$$g_t(G) \geq 0. \tag{43}$$

Case 1: $t \notin \Gamma(S)$. If the set of active constraints is maximal, $g_t = 0$.

Case 2: $t \in \Gamma(S)$. Since each $g_t(G) \in \{0, 1\}$, g_t is the incidence vector of some $\mathcal{P}_t \subseteq \mathcal{G}$: $g_t(G) = 1$ for each $G \in \mathcal{P}_t$ and $g_t(G) = 0$ otherwise. (42) implies that each (i, j) in $I(t, S)$ is contained in at least one $G \in \mathcal{P}_t$. Hence $\mathcal{P}_t$ is a cover. We then show that $\mathcal{P}_t$ is minimal. Suppose not and there exists a cover $\mathcal{P}'$ such that $\mathcal{P}' \subset \mathcal{P}_t$. Then $\mathcal{P}_t \setminus \mathcal{P}' \neq \emptyset$. Then for $\mathcal{P}'$, (42)-(43) continues to hold, but more constraints are active: one more constraint in (43) is active, and since the left-hand side of each constraint in (42) is weakly reduced, more constraints in (42) may become active. But then the set of active constraints at z cannot be maximal and z is not extreme. Hence $\mathcal{P}_t$ is minimal. □

Proof of Theorem 4. By Lemmas 6 and 12, for the extreme rays with given S, it implies that the extreme rays that does not yield the bounds $\bar{c}(I(t, S))$ for all $t \in T$ are redundant and we are done. □

Before the proof of Proposition 1, we introduce the following definition. Let E be a finite set. A function $f : 2^E \to \mathbb{R}$ is called submodular, if $f(U) + f(V) \geq f(U \cap V) + f(U \cup V)$ for all $U, V \subseteq E$; see Fujishige (2005). A function $f : 2^E \to \mathbb{R}$ is nondecreasing, if for all $U, V \subseteq E$, $U \subseteq V$ implies $f(U) \leq f(V)$.

Proof of Proposition 1. We show that f is a nondecreasing submodular function on Ω with $f(\emptyset) = 0$. We first show that $\bar{c}$ is nondecreasing, submodular, and $\bar{c}(\emptyset) = 0$. Since $\mathcal{G}$ is a laminar, by Lemma 1 in Che et al. (2013), we have that $\bar{c}$ is nondecreasing, submodular, and $\bar{c}(\emptyset) = 0$.

Because the operation $I(t, S)$ is in essence the intersection of S and t, for all $S, S' \subseteq \Omega$, we have $I(t, S \cap S') = I(t, S) \cap I(t, S')$ and $I(t, S \cup S') = I(t, S) \cup I(t, S')$. Hence the submodularity of $\bar{c}$ implies that f^t is submodular. Since f is a convex combination of nondecreasing submodular functions (i.e., taking expectation of f^t), f is nondecreasing, submodular, and $f(\emptyset) = 0$. $\square$

Proof of Proposition 2. For any ex post allocation rule p, we can define an equivalent allocation rule $x : T \times \Omega \to \{0, 1\}$ by

$$x^t(t'_i, j) = \begin{cases} p(i, j, t), & \text{if } t'_i = t_i, \\ 0, & \text{otherwise.} \end{cases}$$

The ex post feasibility constraints for p can be written as

$$x^t(S) \leq f^t(S), \text{ for all } t \in T, \text{ all } S \subseteq \Omega. \tag{44}$$

By Proposition 1, $\mathcal{Q}$ is a polymatroid. Let y be an extreme point of $\mathcal{Q}$. By Proposition 6 in Alaei et al. (2019), there exists an ordered subset $\Pi \subseteq \Omega$ such that

$$y(e) = \begin{cases} f(S^k) - f(S^{k-1}), & \text{if } e \in S^k \setminus S^{k-1}, \\ 0, & \text{otherwise,} \end{cases}$$

where S^k is the first k elements of Π. It implies that the interim feasibility condition (22) must be tight for every S^k, which implies that for any ex post allocation rule x that implements y, the ex post feasibility condition (44) must also be tight for every S^k and every $t \in T$. Define the following ordered subset mechanism

$$x^t(t'_i, j) = \begin{cases} 1, & \text{if } t'_i = t_i, (t'_i, j) \in S^k \setminus S^{k-1}, \text{ and } f^t(S^k) \geq f^t(S^{k-1}) + 1, \\ 0, & \text{otherwise.} \end{cases}$$

By construction, x yields the outcome such that the ex post feasibility is tight for every t and S^k, and x implements y. $\square$

Proof of Lemma 7. (1) $(f, g)M(W) = 0$ implies $f(i, t_i) = g(t) = f(j, t_j) = g(t'_i, t_j) = f(i, t'_i)$ for all t_i, t_j, t'_i. Hence $(f, g) = (1, ..., 1)$ is in the linearity space of W and W is non-pointed. (2) is immediate. $\square$

Proof of Lemma 8. Let (f,g) be a $\{0,\pm 1\}$ generator. Since $(f,g) \in W$, for each $t \in T$, we have

$$-f(i,t_i) + g(t) \geq 0. \tag{45}$$

When all of the constraints are active, the generators in the linearity space of W is given by $(1,...,1)$ and $(-1,...,-1)$. We then identify the generators not in the linearity space. We distinguish the following two cases.

(a) $f(i,t_i) \in \{0,-1\}$ for all (i,t_i). Since (f,g) attains a maximal set of active constraints, then $f(i,t_i) = -1$ for a unique (i,t_i) and 0 otherwise, and $g = 0$.

(b) $f(i,t_i) = 1$ for at least one (i,t_i). By (45), for any $t \in T$, if $f(i,t_i) = 1$ for at least one i, then $g(t) = 1$. If $f(i,t_i) = 0$ for exactly one i, and $f(k,t_k) \leq 0$, then $g(t) = 0$. If $f(i,t_i) = -1$ for $i = 1,2$, then $g(t) = -1$. Define $(\tilde{f},\tilde{g})$ by replacing each $f(i,t_i) = -1$ by 0 and $g(t) = -1$ by 0. There are more active constraints at $(\tilde{f},\tilde{g})$ than at (f,g). Hence if (f,g) has a maximal set of active constraints, $f \geq 0$ and (f,g) is the incidence vector of $(E,\Gamma(E))$. $\square$

We introduce the following theorem for the proof of Lemma 9.

Lemma 13. *(Ghouila-Houri, 1962; Schrijver, 1986) Let M be an $p \times q$ matrix. M is totally unimodular if and only if for every subset of columns $\Omega \subseteq \{1,...,q\}$, there exists a partition Ω_1, Ω_2 of Ω such that*

$$\left| \sum_{j \in \Omega_1} m_{ij} - \sum_{j \in \Omega_2} m_{ij} \right| \leq 1 \;\; for \;\; i = 1,...,p. \tag{46}$$

Proof of Lemma 9. A similar argument as Lemma 8 implies that W is non-pointed. We show that the constraint submatrix $M = [\begin{smallmatrix} B \\ C \end{smallmatrix}]$ is totally unimodular. Note that M is given by

$$M = \begin{bmatrix}
1 & 1 & & & & & & \\
& & 1 & 1 & & & & \\
& & & & 1 & 1 & & \\
& & & & & & 1 & 1 \\
1 & & 1 & & & & & \\
& 1 & & 1 & & & & \\
& & & & 1 & & 1 & \\
& & & & & 1 & & 1 \\
1 & & & & 1 & & & \\
& 1 & & & & 1 & & \\
& & 1 & & & & 1 & \\
& & & 1 & & & & 1
\end{bmatrix}.$$

Partition all columns of M into

$$\Omega_1 = \{(a_1, t_{11}), (a_1, t_{22}), (a_2, t_{12}), (a_2, t_{21})\} \text{ and}$$
$$\Omega_2 = \{(a_1, t_{12}), (a_1, t_{21}), (a_2, t_{11}), (a_2, t_{22})\},$$

where t_{kl} denotes the type profile (t_{1k}, t_{2l}). For any subset of the columns Ω, let

$$\Omega_1' = \Omega \bigcap \Omega_1 \text{ and } \Omega_2' = \Omega \bigcap \Omega_2.$$

It can be seen that the two 1s in each row either belongs to different sets Ω_1' and Ω_2', or at least one of the two 1s belongs to neither of Ω_1' and Ω_2'. By the Ghouila-Houri theorem, M is TUM. $\square$

References

Alaei, S., Fu, H., Haghpanah, N., Hartline, J., and Malekian, A. (2019). Efficient computation of optimal auctions via reduced forms. *Mathematics of Operations Research*, 44(3):1058–1086.

Balas, E. (2001). Projection and lifting in combinatorial optimization. *Computational Combinatorial Optimization: Optimal or Provably Near-Optimal Solutions, Juenger, M. and Naddef, D., eds. Springer.*

Balas, E. and Pulleyblank, W. (1983). The perfectly matchable subgraph polytope of a bipartite graph. *Networks*, 13:495–516.

Biró, P., Fleiner, T., Irving, R., and Manlove, D. (2010). The college admissions problem with lower and common quotas. *Theoretical Computer Science*, 411:3136–3153.

Bogomolnaia, A. and Moulin, H. (2001). A new solution to the random assignment problem. *Journal of Economic Theory*, 100(2):295–328.

Border, K. (1991). Implementation of reduced form auctions: A geometric approach. *Econometrica*, 59(4):1175–1187.

Börgers, T. and Postl, P. (2009). Efficient compromising. *Journal of Economic Theory*, 144(5):2057–2076.

Budish, E. and Cantillon, E. (2014). The multi-unit assignment problem: Theory and evidence from course allocation at harvard. *American Economic Review*, 102(5):2237–2271.

Budish, E., Che, Y., Kojima, F., and Milgrom, P. (2013). Designing random allocation mechanisms: Theory and applications. *American Economic Review*, 103(2):585–623.

Cai, Y., Daskalakis, C., and Weinberg, M. (2012). Optimal multi-dimensional mechanism design: Reducing revenue to welfare maximization. *Proceedings of 53rd Annual IEEE Symposium on Foundations of Computer Science*, pages 130–139.

Che, Y., Kim, J., and Mierendorff, K. (2013). Generalized reduced-form auctions: A network-flow approach. *Econometrica*, 81(6):2487–2520.

Cheng, Y. and Yang, Z. (2017). A stable hospital-doctor matching mechanism under distributional and hierarchical constraints. *Preprint*, University of York.

Crawford, V. and Knoer, E. (1981). Job matching with heterogeneous firms and workers. *Econometrica*, 49(1):437–450.

Demange, G. (2004). On group stability in hierarchies and networks. *Journal of Political Economy*, 112:754–778.

Demange, D., G. D. and Sotomayor, M. (1986). Multi-item auctions. *Journal of Political Economy*, 94(1):863–872.

Edmonds, J. (1970). Submodular functions, matroids, and certain polyhedra. *Combinatorial Structures and Their Applications*, Gordon and Breach.

Ehlers, L., Hafalir, I., Yenmez, B., and Yildirim, M. (2014). School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory*, 153:648–683.

Fragiadakis, D. and Troyan, P. (2016). Improving matching under hard distributional constraints. *Theoretical Economics*, 12(2):863–908.

Fujishige, S. (2005). *Submodular Functions and Optimization, 2nd edition*. Elsevier, Amsterdam.

Gershkov, A., Goeree, J., Kushnir, A., Moldovanu, B., and Shi, X. (2013). On the equivalence of bayesian and dominant strategy implementation. *Econometrica*, 81(6):197–220.

Goeree, J. and Kushnir, A. (2016). Reduced-form implementation for environments with value interdependencies. *Games and Economic Behavior*, 99:250–256.

Gopalan, P., Nisan, N., and Roughgarden, T. (2015). Public projects, boolean functions and the borders of border's theorem. *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 395–415.

Hart, S. and Reny, P. (2017). Implementation of reduced form mechanisms: A simple approach and a new characterization. *Economic Theory Bulletin*, 3:1–8.

Hoffman, A. (1976). Total unimodularity and combinatorial theorems. *Linear Algebra and Applications*, 13(1-2):103–108.

Kamada, Y. and Kojima, F. (2015). Efficient matching under distributional constraints: Theory and applications. *American Economic Review*, (1):67–99.

Koopmans, T. and Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica*, 25(1):53–76.

Manelli, A. and Vincent, D. (2010). Bayesian and dominant-strategy implementation in the independent private values model. *Econometrica*, 78(6):1905–1938.

Maskin, E. and Riley, J. (1984). Optimal auctions with risk averse buyers. *Econometrica*, 52(6):1473–1518.

Matthews, S. (1984). On the implementability of reduced form auctions. *Econometrica*, 52(6):1519–1522.

Myerson, R. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73.

Myerson, R. and Satterthwaite, M. (1983). Efficient mechanisms and bilateral trading. *Journal of Economic Theory*, 29(2):265–281.

Schrijver, A. (1986). *The Theory of Linear and Integer Programming*, John Wiley and Sons.

Seymour, P. (1980). Decomposition of regular matroids. *Journal of Combinatorial Theory (B)*, 28(1):305–359.

Seymour, P. (1981). Recognizing graphic matroids. *Combinatorica*, 1(1):75–78.

Shapley, L. and Shubik, M. (1971). The assignment game i: The core. *International Journal of Game Theory*, 1(1):111–130.

Sönmez, T. and Ünver, U. (2010). Course bidding at business school. *International Economic Review*, 51(1):99–123.

Truemper (1985). A decomposition theory for matroids. i. general results. *Journal of Combinatorial Theory, Series B*, 39(1):43–75.

Vohra, R. (2011). *Mechanism Design: A Linear Programming Approach (Econometric Society Monographs)*. Cambridge University Press.

Vohra, R. (2013). Lecture notes in submodularity and discrete convexity. University of Pennsylvania.

Williamson, O. E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*. The Free Press.